Journal of Visualized Experiments

Using eye movements recorded in the visual world paradigm to explore the online processing of spoken language

--Manuscript Draft--

Article Type:	Invited Methods Article - JoVE Produced Video
Manuscript Number:	JoVE58086R4
Full Title:	Using eye movements recorded in the visual world paradigm to explore the online processing of spoken language
Keywords:	Eye tracking technique; visual world paradigm; spoken language; online processing; complex statement; generalized linear mixed model; binomial distribution; familywise error; Bonferroni adjustment
Corresponding Author:	Likan Zhan Beijing Language and Culture University Beijing, Beijing CHINA
Corresponding Author's Institution:	Beijing Language and Culture University
Corresponding Author E-Mail:	zhanlikan@hotmail.com
Order of Authors:	Likan Zhan
Additional Information:	
Question	**Response**
Please indicate whether this article will be Standard Access or Open Access.	Standard Access (US$2,400)
Please indicate the **city, state/province, and country** where this article will be **filmed**. Please do not use abbreviations.	Beijing Language and Culture University, No 15, Xueyuan Rd., Haidian District, Beijing, China

TITLE:

Using Eye Movements Recorded in the Visual World Paradigm to Explore the Online Processing of Spoken Language

AUTHORS AND AFFILIATIONS:

Likan Zhan

Institute for Speech Pathology and the Brain Science, School of Communication Science, Beijing Language and Culture University, Beijing, China

Likan Zhan (zhanlikan@hotmail.com)

KEYWORDS:

Eye tracking technique, visual world paradigm, spoken language, online processing, complex statement, generalized linear mixed model, binomial distribution, familywise error, Bonferroni adjustment

SUMMARY:

The visual world paradigm monitors participants' eye movements in the visual workspace as they are listening to or speaking a spoken language. This paradigm can be used to investigate the online processing of a wide range of psycholinguistic questions, including semantically complex statements, such as disjunctive statements.

ABSTRACT:

In a typical eye tracking study using the visual world paradigm, participants' eye movements to objects or pictures in the visual workspace are recorded via an eye tracker as the participant produces or comprehends a spoken language describing the concurrent visual world. This paradigm has high versatility, as it can be used in a wide range of populations, including those who cannot read and/or who cannot overtly give their behavioral responses, such as preliterate children, elderly adults, and patients. More importantly, the paradigm is extremely sensitive to fine grained manipulations of the speech signal, and it can be used to study the online processing of most topics in language comprehension at multiple levels, such as the fine grained acoustic phonetic features, the properties of words, and the linguistic structures. The protocol described in this article illustrates how a typical visual world eye tracking study is conducted, with an example showing how the online processing of some semantically complex statements can be explored with the visual world paradigm.

INTRODUCTION:

Spoken language is a fast, ongoing information flow, which disappears right away. It is a challenge to experimentally study this temporal, rapidly change speech signal. Eye movements recorded in the visual world paradigm can be used to overcome this challenge. In a typical eye tracking study using the visual world paradigm, participants' eye movements to pictures in a display or to real objects in a visual workspace are monitored as they listen to, or produce, spoken language depicting the contents of the visual world[1-4]. The basic logic, or the linking hypothesis, behind this paradigm is that comprehending or planning an utterance will (overtly or covertly) shift participants' visual attention to a certain object in the visual world. This attention shift will have a high probability to initiate a saccadic eye movement to bring the attended area into the foveal vision. With this paradigm, researchers intend to determine at what temporal point, with respect to some acoustic landmark in the speech signal, a shift in the participant's visual attention occurs, as measured by a saccadic eye movement to an object or a picture in the visual world. When and where saccadic eye movements are launched in relation to the speech signal are then used to deduce the online language processing. The visual world paradigm can be used to study both the spoken language comprehension[1,2] and production[5,6]. This methodological article will focus on comprehension studies. In a comprehension study using the visual world paradigm, participants' eye movements on the visual display are monitored as they listen to the spoken utterances talking about the visual display.

Different eye tracking systems have been designed in history. The simplest, least expensive, and most portable system is just a normal video camera, which records an image of the participant's eyes. Eye movements are then manually coded through frame-by-frame examination of the video recording. However, the sampling rate of such an eye-tracker is relatively low, and the coding procedure is time consuming. Thus, a contemporary commercial eye tracking system normally uses optical sensors measuring the orientation of the eye in its orbit[7-9]. First, to correctly measure the direction of the foveal vision, an infrared illuminator (normally with the wavelength around 780-880 nm) is normally laid along or off the optical axis of the camera, making the image of the pupil distinguishably brighter or darker than the surrounding iris. The image of the pupil and/or of the pupil corneal reflection (normally the first Purkinje image) is then used to calculate the orientation of the eye in its orbit. Second, the gaze location in the visual world is actually contingent not only on the eye orientation with respect to the head but also on the head orientation with respect to the visual world. To accurately infer the gaze of regard from the eye orientation, the light source and the camera of the eye-trackers either are fixed with respect to participants' head (head-mounted eye-trackers) or are fixed with respect to the visual world (table-mounted or remote eye-trackers). Second, the participants' head orientation must either be fixed with respect to the visual world or be computationally compensated if participants' head is free to move. When a remote eye-tracker is used in a head-free-to-move mode, the participants' head position is typically recorded by placing a small sticker on participants' forehead. The head orientation is then computationally subtracted from the eye orientation to retrieve the gaze location in the visual world. Third, a calibration and a validation process are then required to map the orientation of the eye to the gaze of regard in the visual world. In the calibration process, participants' fixation samples from known target points are recorded to map the raw eye data to gaze position in the visual world. In the validation process, participants are presented with the same target points as the calibration process. The difference existing between the computed fixation position from the calibrated results and the actual position of the fixated target in the visual world are then used to judge the accuracy of the calibration. To further reconfirm the accuracy of the mapping process, a drift check is normally applied on each trial, where a single fixation target is presented to participants to measure the difference between the computed fixation positions and the actual position of the current target.

The primary data of a visual world study is a stream of gaze locations in the visual world recorded at the sampling rate of the eye-tracker, ranging over the whole or part of the trial duration. The dependent variable used in a visual world study is typically the proportion of samples that participants' fixations are situated at certain spatial region in the visual world across a certain time window. To analyze the data, a time window has firstly to be selected, often referred to as periods of interest. The time window is typically time-locked to the presentation of some linguistic events in the auditory input. Furthermore, the visual world is also needed to split into several regions of interest (ROIs), each of which is associated with one or more objects. One such region contains the object corresponding to the correct comprehension of the spoken language, and thus it is often called the target area. A typical way to visualize the data is a proportion-of-fixation plot, where at each bin in a time window, the proportion of samples with a look to each region of interest are averaged across participants and items. Using the data obtained from a visual world study, different research questions can be answered: a) On the coarse-grain level, are participants' eye movements in the visual world affected by different auditory linguistic input?; b) If there is an effect, what is the trajectory of the effect over the course of the trial? Is it a linear effect or high-order effect?; and c) If there is an effect, then on the fine-grain level, when is the earliest temporal point where such effect emerges and how long does this effect last?

To statistically analyze the results, the following points should be considered. First, the response variable, *i.e*., proportions of fixations, is both below and above bounded (between 0 and 1), which will follow a multinomial distribution rather than a normal distribution. Henceforth, traditional statistical methods based on normal distribution such as t-test, ANOVA, and linear (mixed-effect) models[10], cannot be directly utilized until the proportions have been transformed to unbounded variables such as with empirical logit formula[11] or have been replaced with unbounded dependent variables such as Euclidean distance[12]. Statistical techniques that do not require the assumption of normal distribution such generalized linear (mixed-effect) models[13] can also be used. Second, to explore the changing trajectory of the observed effect, a variable denoting the time-series has to be added into the model. The variable is originally an eye-tracker's sampling points realigned to the onset of the language input. Since the changing trajectory is typically not linear, a high-order polynomial function of the time-series is normally added into a (generalized) linear (mixed-effect) model, *i.e*., growth curve analyses[14]. Furthermore, participants' eye positions in the current sampling point is highly dependent on the previous sampling point, especially when the recording frequency is high, resulting in the problem of autocorrelation. To reduce the autocorrelation between the adjacent sampling points, the original data are often down-sampled or binned. In recent years, the generalized additive mixed effect models (GAMM) have also been used to tackle the autocorrelated errors[12,15,16]. The width of bins varies among different studies, ranging from several milliseconds to several hundred milliseconds. The narrowest bin a study can use is restricted by the sampling rate of the eye tracker used in the study. For example, if an eye tracker has a sampling rate of 500 Hz, then the width of the time window cannot be smaller than 2 ms = 1000/500. As we described earlier, the trajectory analysis informs the researcher whether the effect observed on the coarse-grain level is linear with respect to the changing of the time, but the trajectory analysis does not show when the observed effect begins to emerge and how long the observed effect lasts. To determine the temporal position where the observed difference starts to diverge, and to the duration of the temporal period that the observed effect lasts, a statistic analysis has to be repeatedly applied to each time bin. Third, when a statistical analysis is repeatedly applied to each time bin of the periods of interest, the familywise error induced from these multiple comparisons should be tackled. The familywise error will not disappear as long as multiple comparisons have been conducted, no matter what the statistical method is. This familywise error is traditionally corrected with Bonferroni adjustment[17]. Recently, a method called nonparametric permutation test originally used in neuroimaging filed[18] has been applied to the visual word paradigm[19] to control for the familywise error.

Researchers using the visual world paradigm intend to infer the comprehension of some spoken language based on their eye movements in the visual world. To ensure the validity of this deduction, other factors that could influence participants' eye movements should be ruled out or be controlled. The following two factors are among the common ones that need to be considered. The first factor involves some systematic patterns in participants' explanatory fixations, independent of the language input, such as the tendency to fixate on the top left quadrat of the visual world, and moving eyes in the horizontal direction being easier than in the vertical direction, *etc*.[12,20] To make sure that the observed fixation patterns are related to the objects, not to the spatial locations, the spatial positions of different objects should be counterbalanced across different trials or across different participants. The second factor is the basic image features of the objects in the visual world, such as luminance contrast, color and edge orientation, among others[21]. To diagnose this potential confounding, the visual display is normally presented prior to the onset of the spoken language or prior to the onset of the critical acoustic marker of the spoken language, for about 1000 ms. As the language input or the disambiguation point of the language input has not been heard yet, any difference between different conditions observed during this period should be deduced to other confounding factors such as the visual display per se, rather than the language input. Henceforth, eye movements observed in this preview period provide a baseline for determining the effect of the linguistic input. This preview period also allows the participants to get familiarized with the visual display, and to reduce the systematic bias of the explanatory fixations when the spoken language is presented.

To illustrate how a typical eye tracking study using the visual world paradigm is conducted, the following protocol describes an experiment adapted from L. Zhan[17] to explore the online processing of semantically complex statements, *i.e.,* disjunctive statements (*S1 or S2*), conjunctive statements (*S1 and S2*), and *but*-statements (*S1 but not-S2*). In ordinary conservation, the information expressed by some utterances is actually stronger than its literal meaning. Disjunctive statements like *Xiaoming's box contains a cow or a rooster* are such utterances. Logically, the disjunctive statement is true as long as the two disjuncts *Xiaoming's box contains a cow* and *Xiaoming's box contains a rooster* are not both false. Therefore, the disjunctive statement is true when the two disjuncts are both true, where the corresponding conjunctive statement *Xiaoming's box contains a cow and a rooster* is also true. In ordinary conversation, however, hearing the disjunctive statement often suggests that the corresponding conjunctive statement is false (*scalar implicature*); and suggests that the truth values of the two disjuncts are unknown by the speaker (*ignorance inference*). Accounts in the literature differ in whether two inferences are grammatical or pragmatic processes[22-26]. The experiment shows how the visual world paradigm can be used to adjudicate between these accounts, by exploring the online processing of three complex statements.

PROTOCOL:

All subjects must give informed written consent before the administration of the experimental protocols. All procedures, consent forms, and the experimental protocol were approved by the Research Ethics Committee of the Beijing Language and Culture University.

NOTE: A comprehension study using the visual world paradigm normally consists of the following steps: Introduce the theoretical problems to be explored; Form an experimental design; Prepare the visual and auditory stimuli; Frame the theoretical problem with regard to the experimental design; Select an eye-tracker to track participants' eye movements; Select a software and build a script with the software to present the stimuli; Code and analyze the recorded eye-movements data. A specific experiment can differ from each other in any of the described steps. As an example, a protocol is introduced to conduct the experiment and discuss some points that researchers need to keep in mind when they build and conduct their own experiment using the visual world paradigm.

1. Prepare Test Stimuli

1.1. Visual stimuli

1.1.1. Download 60 clip arts of animals that are free of copyright from the internet. Open each image with an image editor (e.g., Pixelmator), click **Tools | Quick selection tool** to select and delete the background. Click **Image | Image Size** to resize them to 120 × 120 pixels.

1.1.2. Invite a student majoring in painting to draw four light green boxes, as illustrated in **Figure 1**. Use the image editor to rescale the big open box to 320 × 240 pixels, the small closed box with the size of 160 × 160 pixels, and the two small open boxes to 160 × 240 pixels, respectively.

1.1.3. Click **Pixelmator | File | New** to build a template of the test image with the size of 1024 × 768 pixels. Drag the animals and the boxes to the correction locations illustrated in **Figure 1**.

NOTE: The layout of the test image varies between studies, but the optimal way is to use four objects and to put them at the four quadrants of the test image. In this way, it is easier to counterbalance the spatial position of the objects.

1.1.4. Create 60 test images like **Figure 1**, with each animal image being used twice. Counterbalance the spatial locations of the four boxes among the images.

NOTE: The number of the images does not have to be exact 60, as long as their effect is dissociable from that of the experimental manipulations.

1.2. Spoken language stimuli

1.2.1. Design four test sentences corresponding to each test image and 240 test sentences in total to be recorded. Ensure that three of the four sentences are in the form of **Figure 2**; and the filler sentence is in the form of *Xiaoming's box doesn't contain a rooster but a cow*.

NOTE: The test sentences should be presented in the native language that participants speak. The participants in this experiment are Chinese from Beijing, Mainland China, so the test language is Mandarin Chinese.

1.2.2. Recruit a female native speaker (a native speaker of Mandarin Chinese in this experiment) to record four example statements like **Figure 2**, as well as audio of all the animals being used in the experiment. When recording the isolated animal names, ask the speaker to imagine that the names of the animals are intact components of a simple sentence, such as *Xiaoming's box contains a/an* ___, but she only needs to pronounce the name of the animal overtly.

1.2.3. Replace the audio segments of the two animals in the example statements with the audio of the two animals used in each trial to create the full list of the test audios. First, open Praat (Any other audio editing software is an eligible alternative) and click **Open | Read from file | Navigate to the file | Open and edit***,* navigate to an element to be replaced, and click **View and Edit | Edit | Copy selection to sound clipboard***.* Second, use the same steps to open an example statement, click **paste after selection**. Third, click **Save | save as wav file** to save the edited statement. Repeat the process for all the elements to be changed and all the test sentences.

1.2.4. Recruit about 10 native speakers of the test language (Mandarin Chinese here) to determine whether or not the constructed test audio is intelligible and natural.

NOTE: The test audio is traditionally recorded as a whole, rather than as separate words. This traditional recording method is reasonable if the test audio are themselves separate words. If the spoken language stimuli are sentences rather than separate words, however, this traditional method has several shortcomings: First, a ubiquitous property of a continuous speech is that two or more speech sounds tend to temporally and spatially overlap, which makes it hard to determine the onset of the critical word. Second, the variance between the length of different trials also makes it difficult to combine all the trials together for statistical analyses. Third, this method is often time consuming especially when the numbers of the test audio are relatively large, such as the experiments we reported in the protocol. To overcome the shortcomings of the traditional recording method, a different method is proposed to construct the spoken test audios. First, a list of sample sentences containing the words that are common among all the test audio was recorded. Second, all the words that change between trials in isolation were also recorded. Finally, the sample sentences were replaced with the recorded words to construct the full list of the test audios. Compared to the traditional method, this method has several advantages. First, all the test audio is exactly the same except for the critical words, and all the potential confounding effects in the test audio are henceforth controlled. Second, being the same in length also makes the segmentation of the test audio easier than when the test audio is recorded as a whole. One potential disadvantage of this method is that the constructed audio might be not natural. Henceforth, the naturalness of the test audio has to be evaluated before they are eligible for the actual testing.

1.3. Divide the 240 test sentences into four groups, with each group containing 15 conjunctive statements, 15 disjunctive statements, 15 *but* statements, and 15 filler sentences. Ensure that each participant encounters only one group of 240 trials: he/she sees all the test images but hears only one group of the test audios.

NOTE: This is to address the concern that if the same stimulus is repeated, participants might be getting accustomed to these stimuli and possibly even becoming strategic about how participants have responded to the stimuli.

1.4. Save all important information regarding the test stimuli into a tab-delimited txt file, with each row corresponding to each of the 240 trials. Ensure that the file includes at least the following columns: *experiment_group*, *sentential_connective, trial_number, test_image, test_audio, test_audio_length, ia_top_left, ia_top_right, ia_bottom_left, ia_bottom_right, animal_1_image*, *animal_1_audio*, *animal_1_audio_length*, *animal_2_image*, *animal_2_audio*, *animal_2_audio_length*.

NOTE: Experiment_group is used to split the 240 trials into 4 groups. Sentential_connective corresponds to different experimental conditions. Animal_1_image corresponds to the image of the animal that will be firstly presented to familiarize the participants with the animals used in the test image. Test_image, test_audio, and test_audio_length refer to the test image and the test audio as well its length used in the current trial. ia_top_left, ia_top_right, ia_bottom_left, ia_bottom_right refer to the name of the four interest areas in the current trial, *i.e.*, whether it is a “Big open” box, “small closed” box, the small open box containing the “first mentioned” animal in the test audio, or the small open box containing the “second mentioned” animal in the test audio. Animal_1_audio and animal_1_audio_length refer to the audio and length of the audio corresponding to the animal_1_image. Animal_2_image, animal_2_audio, and animal_2_audio_length correspond the second animal that will be presented. One thing to stress is that the sequence to present the two animals is counterbalanced with respect to whether the animal is mentioned in the first or in the second half of the test audios.

2. **Frame the Theoretical Prediction with regard to the Experimental Design.**

2.1. Ensure participants’ behavioral responses and eye-movements in the experimental design can be used to differentiate comprehensions of the test sentences and can be used to adjudicate between different accounts to be tested.

NOTE: Given the experimental design, the correct response to a conjunctive statement is the big open box, such as Box A in **Figure 1**. The correct response to a *but*-statement is the small open box containing the animal being mentioned in the first half of the test audios, such as Box D in

Figure 1. Participants' responses to the disjunctive statement, however, depend on whether and/or how the two discussed inferences are processed. If participants compute neither the scalar implicature nor the ignorance inference, then all the four boxes are eligible options. If participants compute the scalar implicature but not the ignorance inference, then the big open, such as box A in **Figure 1**, will be ruled out, and the remaining three boxes B, C, and D are all eligible options. If participants compute the ignorance inference but not the scalar implicature, then the small open box(es) will be ruled out, *i.e.*, boxes C and D will be ruled out. To summarize, the small closed box, such as box B in **Figure 1**, will not be chosen as the final option of a disjunctive statement until the scalar implicature and the ignorance inferences are both computed.

3. Build the Experimental Script

3.1. Open the **Experiment Builder***,* click **File | New** to create an experiment project. Input the project name such as *vwp_disjunction*. Select the project location. Check *EyeLink Experiment* and choose *Eyelink 1000plus* from the drop list. These operations will create a subdirectory containing all files related to the experiment It will create a subdirectory named *vwp_disjunction* with a file named "graph.ebd" in the folder.

NOTE: Experiment Builder is used to build the experimental script to present the test stimuli and to record participants' eye movements as well as their behavioral responses. The Experiment Builder is a What-You-See-Is-What-You-Get tool to build experimental script. It is easy to use, but any other stimuli presentation software is an eligible alternative.

3.2. Visualize the hierarchical structure of a typical eye-tracking experiment using the visual world paradigm as seen in **Figure 3**. Each pink rectangle in the figure is implemented as a SEQUENCE object by Experiment Builder; and each object with gray background is implemented as a node object.

NOTE: A SEQUENCE in the Experiment Builder is an experimental loop controller used to chain together different objects as a complex node. A sequence always begins with a *START* node. And a data source can be attached to a sequence node to supply different parameters for each trial.

3.3. Build the Experiment sequence

3.3.1. Click **File | Open**, browse to the directory of experiment and double click the *graph.ebd* file in the project directory to open the saved experiment project.

3.3.2. Click **Edit | Library Manager | Image | Add** to load the images into the experiment Project. Similarly, click **Edit | Library Manager | Sound | Add** to load the audio into the experiment project.

3.3.3. Drag a DISPLAY_SCREEN object into the work space and change its label value on the properties panel to rename it as **Instruction**. Double click to open the **Instruction** node, and click the **Insert Multiline Text Resource** button to input the experimental instruction. Ensure the instruction contains the following information:

In each trial, first you will see images of two animals, one animal each printed on the screen in turn, along with the audio of the animals played on the two speakers situated at both sides of the screen. A black dot will then be presented at the center of the screen. You should press the SPACE key while fixating on the dot. Next, you will see a test image consisting of four boxes printed on the screen and hear a test sentence being played via the two speakers. Your task is to locate Xiaoming's box according to the test sentence you heard and press the corresponding button as soon as possible:

Top left box --- Left arrow
Top Right Box --- Up arrow
Bottom left box --- Left arrow
Bottom right box --- Right arrow

In each test image, you will see four boxes situated at the four quadrants and two animals containing in the boxes. The four boxes can vary in two dimensions: its closeness and its size. Whether a box is closed or not influences our epistemic knowledge on that box, but not the animal(s) it contains. If a box is open, then the animal(s) contained in that box is known. If a box is closed, then the animal(s) contained in that box is unknown. The size of a box affects the number of animals contained in the box, but not our epistemic knowledge on that box. No matter the box is closed or not, a small box only and always contains one animal, and a big box always contains two different animals.

If you are comfortable with the experimental aim and the procedure, please let the experimenter know and we will help you to perform the standard eye tracking calibration and validation routines. If you have any questions, please don't hesitate to ask.

Note: This is an instruction that will be printed on the screen prior to the experiment (The instructions should be written in the native language the participants speak, such as Mandarin Chinese here).

3.3.4. Drag a KEYBOARD object into the work space.

NOTE: This step is used to end the *Instruction* screen

3.3.5. Drag a *SEQUENCE* object into the work space and rename it as *Block*.

3.3.6. Select the *Block* sequence, click the value field of the **Data Source** property to bring up the **Data Source Editor**. Click the **Import Data** Button on the data source editor screen, brow to the .txt file created in step 1.4 to import the data source.

3.3.7. Click the **Randomization Setting** button in the data source editor, check **Enable Trial**

Randomization, select *trial_number* from the value field of the *Column* field, and select *experimental_group* from the drop-list of the *Splitting Column* field.

3.3.8. Drag the second *DISPLAY_SCREEN object to the work space and rename it as Goodbye.* Double click the goodbye node and insert the following information: *The experiment is finished and Thank you for your participation.*

3.3.9. Left-click on the *START* node, drag the arrow to the *Instruction* node, and release the mouse button to connect the START node to the *Instruction* node. Repeat the same mouse moves to connect *Instruction* node to the *KEYBOARD* node, KEYBOARD node to *Block* node, then *Block* node to the *Goodbye* node. Click **View | Arrange Layout** to arrange the nodes in the workspace.

3.4. Build the block sequence

3.4.1. Double click to open the **Block** sequence. Drag an **El_CAMERA_SETUP** node into the **Block** sequence to bring up a camera setup screen on the EyeLink Host PC for the experimenter to perform camera setup, calibration, and validation. Click the **Calibration Type field** in the Properties panel and choose HV5 from the dropdown list.

NOTE: The number of locations in the mapping process varies between different experimental designs. The more locations sampled and the more space covered, the greater the accuracy can be expected. But more samples mean more time to finish the processes. So practically, the number of locations in a specific study cannot be very big, especially when the participants are preliterate children or clinical patients. In the visual world paradigm, the number of the interest areas is relatively small, and the areas of interest are normally relatively big. The mapping process can reach a satisfying level with relatively small number of locations. In the protocol we described, we used a five points' calibration and validation.

3.4.2. Drag a SEQUENCE node into the *Block* sequence and rename it as Trial. Connect the START node to the CAMERA_SETUP node, then to the SEQUENCE node.

3.5. Build the **Trial** sequence

3.5.1. Drag a DISPLAY_SCREEN node into the **Trial** sequence and rename it as *animal_1_image*. Double click to open the *Screen Builder* node and click the *Insert Image Resource* button on the Screen Builder toolbar to insert an animal image from the uploaded image sources. Click the value field of the *Source File Name property*, navigate to the *DataSource* attached to the *Block* Sequence; and double click the *Animal_1_Image* column to connect the DISPLAY_SCREEN with correct column of the data source.

3.5.2. Drag a PLAY_SOUND node into the **Trial** sequence and rename it as *animal_1_audio*. Click the *Sound File* property of the *animal_1_audio* node and connect it with the correct column of the data source (as being described in step 3.5.1).

3.5.3. Drag a TIMER node into the **Trial** sequence and rename it as *animal_1_audio_length*. Click the *Duration* Property of the TIMER node and navigate to the correct column of the data source created in 3.4.1.

3.5.4. Drag another **DISPLAY_SCREEN** node, another **PLAY_SOUND** node, and another **TIMER** node into the Trial sequence, rename them as *Animal_image_two* an, *Animal_audio_two,* and *Audio_duration_two,* and repeat the steps being described in steps 3.5.1 - 3.5.3.

NOTE: These steps are included to control for the potential confounding that the same image might be named differently by different participants. Counterbalance the sequence of presenting the two animals with respect to whether it is mentioned in the first or second half of the test audios.

3.5.5. Drag a **Prepare Sequence** object into the Trial Sequence and change the property **Draw To Eyelink Host to IMAGE**.

NOTE: This node is used to preload the image and audio files to memory for real-time image drawing and sound playing. And it is also used to draw feedback graphics on the Host PC so that the participants' gaze accuracy can be monitored.

3.5.6. Drag a **DRIFT_CORRECT** node into the **Trial** sequence to introduce the drift correction.

3.5.7. Drag a new SEQUENCE node and rename it as *Recording*. Connect the START to these nodes one after one.

3.6. Build the **Recording** sequence

3.6.1. Check the Record field, and double click to open the *Record* sequence.

NOTE: A sequence with *Record* property checked means that participants' eye movements during this period will be recorded.

3.6.2. Drag a new **DISPLAY_SCREEN** into the **Record** sequence, rename it as *Test_Image*. Add the message *Test_Image_Onset* into the *Message* property of the *Test_Image* node.

3.6.3. Double click to open the *Screen Builder* node and click the *Insert Image Resource* button on the Screen Builder toolbar to insert an animal image from the uploaded image sources. Click the value field of the *Source File Name property*, navigate to the *DataSource* attached to the *Block* Sequence; and double click the *Test_Image* column to connect the DISPLAY_SCREEN with correct column of the data source.

3.6.4. Double click the DISPLAY_SCREEN node to open the **Screen Builder**, click the button of **Insert Rectangle Interest Area Region**, and draw four rectangular areas of interest as illustrated by the blue boxes in **Figure 1**. Change the labels of the four areas of interest to *Top_Left*, *Top_Right*, *Bottom_Left*, and *Bottom_Right*, and connect the *DataViewer Name* filed with the correct columns of the data source.

NOTE: These areas are invisible to the participants. To make the areas of interest more meaningful, label the name of top left area in the example as “Box A (big open)”, area top right area as “Box B (small closed)”, bottom left area as “Box C (second mentioned)”, and area bottom right area as “Box D (First Mentioned)”, because the two small open boxes contain the two animals being mentioned in the first and second half of the test audios, respectively.

3.6.5. Drag a **TIMER** node into the workspace, rename it as *Pause*, and change the **Duration** property to 500 ms.

NOTE: This TIMER node adds some time lag between the onset of the test image and the onset of the test audio and add a PLAY_SOUND node to play the test audios. The time lag gives participants a chance to familiarize with the test images. Participants’ eye movements during this preview period provide a baseline for determining the effects of the spoken language input, especially when the critical words are situated at the beginning of the test audios.

3.6.6. Drag a **PLAY_SOUND** node in to the work space and rename it as *test_audio*. Click the **Sound File** property and connect it with the correct column of the data source (as being described in step 3.5.1) and add the message *Test_Audio_Onset* into the *Message* property.

3.6.7. Drag a **TIMER** node into the work space, rename it as *test_audio_length*. Change the **Duration** Property to 10500 ms.

3.6.8. Add a new **TIMER** node, rename it as *record_extension*, and change the **Duration** property to 4000 ms.

3.6.9. Add a new KEYBOARD node into the work space, rename it as *behavioral responses*, and change the acceptable **Keys** property to “[Up, Down, Right, Left]”.

NOTE: Participants’ behavioral choices can be used to double check the validity of the conclusion deduced from participants’ eye movements.

3.6.10. Connect the **START** node to *Pause*, *test_audio*, *test_audio_length*, then to *Record_extension* node. Add another connection from *test_audio_length* to *behavioral_responses* node.

NOTE: By adding these connections, a new trial will be started after participants made a key press to choose *Xiaoming*’s Box, or 4000 ms after the offset of the test audio.

3.6.11. Drag a **VARIABLE** node into the work space, rename it as *key_pressed*, and connect its value to **behavioral_Responses Keyboard | Triggered Data |** *Key*.

3.6.12. Drag a **RESULT_FILE** node into the work space, drag an **ADD_TO_RESULT_FILE** node into the work space, and connect both the *record_extension* node and the *behavioral_responses* node to the **ADD_TO_RESULT_FILE** node.

3.7. Click **Experiment | Build** to build the experimental script, click **Experiment | Test** run to test run the experiment. After everything is done, click **Experiment | Deploy** to create an executable version of the experimental project.

NOTE: For more information on how to use the Experiment Builder, please consult the software manual [27].

4. Recruit Participants

4.1. Ensure the participants to have normal or corrected normal vision. Recommend that the short-sighted participants to wear contact lenses, but glasses are also acceptable as long as the lenses are clean. Ensure that all participants are native speakers of the testing language, such as Mandarin Chinese here.

NOTE: As a general guideline, a participant is regarded as eligible as long as the participant can see the test images at a distance of about 60 centimeters. In terms of the number of participants, according to some rules of thumb, the number of participants for regression analysis should be no less than 50. Here, thirty-seven postgraduate students from the Beijing Language and Culture University participated in the experiment, which is a little smaller than the recommended amount.

5. Conduct the Experiment

5.1. Select an eye tracker to record participants' eye movements.

NOTE: The eye tracker used in this experiment is Eyelink 1000plus running under the free-to-move head mode. This is a video-based, desktop mounted eye tracking system, using the principle of pupil with corneal reflection (CR) to track eye's rotation. When running under the free-to-move head mode, the eye tracker has the monocular sampling rate of 500 Hz, with a spatial resolution of 0.01° and an average error of less than 0.5°. For more detailed information of the system, please consult its technical specification[28,29]. Alternative trackers can be used, but the ones with remote tracking mode are better, especially when participants are preliterate children.

5.2. Boot the system on the Host PC to start the Host application of the camera.

5.3. To configure the system to desktop remote mode, click the **Set Option** button, set the **Configuration** option to *Desktop -- Target Sticker -- Monocular -- 16/25mm length -- RTARBLER.*

5.4. Click the executable version of the experimental project on the Display PC, input participant's name, and choose a group from the Prompt window: **Select Condition Value to Run**.

NOTE: Each test session will create a folder with the inputted name under the subdirectory *Results* of the experiment project. The EDF file under the folder contained relevant eye movements data.

5.5. Ask the participants to sit approximately 60 cm from a 21 inch, 4:3 color monitor with 1024px x 769px resolution, where 27 pixels equals to 1 degree of angle.

5.6. Adjust the height of the Display PC monitor, to ensure that when the participant is seated and looking straight ahead, they are looking vertically at the middle to top 75% of the monitor.

NOTE: The chair, desk, and/or the PC monitor are preferred if they are adjustable in height. The chair and the desk with casters should be avoided, as they tend to cause unintentional move and roll.

5.7. Place a small target sticker on participants' forehead, to track the head position even when the pupil image is lost, such as during blinks or sudden movements.

NOTE: Different eye trackers might use different methods to track participants' head. To maximize the lateral movement range of the subject, the tracked eye should be on the same side as the illuminator.

5.8. Rotate the focusing arm on the desk mount to bring the eye image into focus.

5.9. Click the **Calibrate** button on the host PC to conduct the calibration process by asking participants to fixate a grid of five fixation targets in random succession with no overt behavioral responses, to map participants' eye movements to the gaze of regard in the visual world.

5.10. Click the **Validate** button on the host PC to validate the calibrated results by asking participants to fixate the same grid of fixation targets. Repeat the calibration and validation routines, when the error is bigger than 1°.

5.11. Conduct the two routines at the beginning of the experiment and whenever the measurement accuracy is poor (*e.g.*, after strong head movements or a change in the participants' posture).

5.12. Click the **Record** button on the host PC to start the experiment.

5.13. Perform a drift check on each trial by asking participants to press the **SPACE** key on the keyboard while fixating at the black dot presented in the center of the screen.

NOTE: When the participants are preliterate children or clinical patients, explicitly instructing them to press the keyboard while fixating the black dot is normally impractical. But their attention and eye fixations tend to be automatically attracted by the displayed black dot. In this case, the experimenter should be the person to press the keyboard while the participant is fixating on the black dot.

5.14. Present the visual stimuli via the Display PC monitor and play the auditory stimuli via a pair of external speakers situated to the left and right of the monitor (earphones are also acceptable).

NOTE: The recordings are played from the hard disk as 24 kHz mono sound clips. If there is no special reason, mono sound clips are preferred to stereo sound clips. In a stereo sound clip, the difference between the two sound tracks, as well as the difference between the two speakers might affect participants' eye movements. For more information on how to use the eye tracker, please consult the user manual[30].

6. Data Coding and Analyses

6.1. Open Data Viewer, click **File | Import File | Import Multiple Eyelink Data Files** to import all the recorded eye tracker files (with the extension of EDF), and save them into a single *.EVS* file.

6.2. Open the saved EVS file and click **Analysis | Reports | Sample Report** to export the raw sample data with no aggregation.

NOTE: If the eye tracker has a sampling rate of 500 Hz, the exported data will have 500 data points, henceforth 500 rows, per second per trial. If participants' left eye is tracked, ensure the following columns as well as the variables created in the data source are exported: RECORDING_SESSION_LABEL, LEFT_GAZE_X, LEFT_GAZE_Y, LEFT_INTEREST_AREA_LABEL, LEFT_IN_BLINK, LEFT_IN_SACCADE, LEFT_PUPIL_SIZE, SAMPLE_INDEX, SAMPLE_MESSAGE. For more information on how to use the Data Viewer, please consult the software manual[31].

6.3. Restrict the statistical analyses to the temporal window from the onset of the test image to the offset of the test audios, *i.e.*, the temporal window with the duration of 11 s.

6.4. Delete the samples where participants' eye movements are not recorded, such as participants blink their eyes, which roughly affects 10% of the recorded data.

NOTE: This is an optional step, as the results are normally the same no matter whether these samples deleted.

6.5. Code the data. To construct the data for a specific area of interest in a certain sampling point, code the data as 1 if participants' eye fixation is situated in the area of interest to be analyzed at that sampling point. Code the data as 0 if the eye fixation is not situated in the areas of interest at that sampling point.

6.6. Draw a proportion-of-fixation to visualize the obtained data. To calculate the proportion-of-fixations over certain area of interest, average the coded data for all the trials and for all the participants in each sample point under each condition. Plot the calculated proportion-of-fixations on the y-axis against the sampling point on the axis, with different panels denoting areas of interest and with the plotting colors denoting different experimental conditions.

NOTE: In the experiment, the four panels depicted participants' fixation patterns on the four areas of interest. The red, green, and blue lines illustrated participants' fixation patterns when the test statements were conjunctions (*S1 and S2*), but-statements (*S1 but not S2*), and disjunctions (*S1 or S2*), respectively. The software used to draw the descriptive plot is the ggplot2 package from R environment. Other software is also available. **Figure 5** is an example of such plot.

6.7. Fit a binomial generalized linear mixed model (GLMM) on each area of interest at each sampling point, as the data was coded as either 1 or 0, depending on whether the participant's fixation is situated in or out of the area of interest at that sampling point.

NOTE: As the data is not binned, and the coded data can only be 1 or 0, so the distribution of the coded data is binary rather than normal. Henceforth, use a GLMM model with the family of binomial distribution. The GLMM model includes a fixed term, the experimental conditions, and two random terms, participants and items. The formula evaluated to the two random terms includes both the intercepts and the slope of the experimental conditions. The software used to do the model fitting is the *lme4* package from R environment. Other software is also available. One thing should be mentioned is that the baseline of the fixed items differed when the analyzed interest area, *i.e.*, the analyzed boxes, are different. To be specific, the conjunction (*S1 and S2*) was chosen as the baseline when analyzing the big-open box (Box A), the disjunction (*A and B*) was chosen as the baseline when analyzing the small-closed box (Box B), and the but-statement was chosen as the baseline when analyzing the first-mentioned box (Box D).

6.8. Bonferroni adjust the *p* values obtained with Wald *z* test, to reduce the familywise error induced by multiple comparisons.

NOTE: Bonferroni adjustment is the traditional way to tackle the familywise error induced by multiple comparisons. Other methods are also available, as we described in the introduction section.

REPRESENTATIVE RESULTS:

Participants' behavioral responses are summarized in **Figure 4**. As we described earlier, the correct response to a conjunctive statement (*S1 and S2*) is the big open box, such as Box A in **Figure 1**. The correct response to a *but*-statement (*S1 but not S2*) is the small open box containing the first mentioned animal, such as Box D in **Figure 1**. Critically, which box is chosen to the conjunctive statements (*S1 or S2*) depends on how a conditional statement is processed: The small closed box, such as box B in Figure 1, is chosen only when the scalar implicature and the ignorance inferences are both computed. **Figure 4** illustrates that when comprehending a disjunctive statement (*S1 or S2*), participants compute both the two inferences.

Furthermore, participants' eye fixations on the small closed box, especially when these fixations are followed by the behavioral responses on that box, also suggest that the scalar implicature and the ignorance inferences are computed. The two inferences should have been processed no later than that of the earliest temporal point when participants shift their visual attention and fixations on the small closed box. Participants' eye-movements are summarized in **Figure 5**. As we can see in panel B, participants' eye-fixations on the small-closed box (box B) don't increase unless the sentential connective is the disjunctive connective, *or*. Furthermore, this increase begins no later than the offset of the disjunctive connective. This suggests that both the scalar implicature and the ignorance inferences are computed before the offset of the sentential connective, i.e., immediately upon encountering the disjunctive connective.

FIGURE AND TABLE LEGENDS:

Figure 1. An example of the test images used in the experiment. The gray area is the test image that is actually presented to the participants. The blues boxes, the dotted lines, and the pixels denoting the width of the elements are only for the purposes of illustration and are invisible to the participants. (Adapted from experiment one of L. Zhan [17] with permission).

Figure 2. An example of the test sentences used in the experiment (Reprinted from L. Zhan [17] with permission).

Figure 3. The hierarchical structure of a typical eye-tracking experiment using the visual world paradigm. The audio illustrated in the image are the English translations of the Mandarin Chinese used in the experiment.

Figure 4. Participants' behavioral responses recorded in the experiment (Adapted from experiment one of L. Zhan [17] with permission).

Figure 5. Participants' eye movements observed in the experiment. The onset and offset of the sentential connectives are signified by two dashed vertical lines. A significant difference existing between the disjunctive and the baseline condition at certain sampling point is signified by the gray area ($p < .05$, Bonferroni adjusted) (Adapted from L. Zhan [17] with permission).

DISCUSSION:

To conduct a visual world study, there are several critical steps to follow. First, researchers intend to deduce the interpretation of the auditorily presented language via participants' eye movements in the visual world. Henceforth, in designing the layout of the visual stimuli, the properties of eye movements in a natural task that potentially affect participants' eye movements should be controlled. The effect of the spoken language on participants' eye movements can then be recognized. Second, acoustic cues in the spoken language are transient and there are no acoustic signatures solely corresponding to certain linguistic category. To correctly time-lock participants' eye-movements with the onset of some linguistic marker, researchers should find an objective and consistent way to define the boundary of some linguistic structure. Third, to correctly map participants' eye rotation around the head with their gaze of regard in the visual world, researchers need to perform one or several runs of calibration, validation, and drift correction processes. Fourth, the data obtained from a visual world study have some peculiar properties, such as being lower and upper bounded, and having autocorrelated errors etc. These peculiar properties should be considered when a method is selected to statistically analyze the data.

A visual world study consists of three essential components: visual display, spoken language, experimental task. Modifications can be made on any component to fulfill researchers' specific purposes. First, a visual display is normally a screening display depicting an array of pictures. But it can also be a screening display depicting an array of printed words[32], a schematic scene[30,31], or a real world scene containing real objects[1,32]. Second, the spoken utterances can be a word[36], a simple sentence[30,31], a semantically complex statement[17,34,35], or a dialogue[39]. Third, in terms of the experimental task, participants are either simply being asked to look at the visual world and listen carefully to the auditory utterances[30,31]; or are being required to make some behavioral responses, such as acting out the movements described by the auditory utterance[1], determining whether or not the auditory utterance applies to the visual display[38], or choosing the correct image in the visual display the spoken utterance is talking about[17].

The visual world paradigm, compared with other psycholinguistic techniques, has several unique advantages. First, the visual world paradigm can be used in a wide of populations, including those who cannot read and/or who cannot overtly give their behavioral responses, such as preliterate children[37-42], elderly adults[46], and patients (*e.g.*, aphasics)[47]. This is because the language stimuli are presented in the auditory rather than in the form of written texts; and the language comprehension is inferred from participants' implicit automatic eye movements rather than from their overt behavioral responses. Second, the visual world paradigm is extremely sensitive to fine grained manipulations of the speech signal. This paradigm can be used to study the online processing of most topics in language comprehension at multiple levels, such as the fine grained acoustic phonetic features[33,45,46], the properties of words[30,31], the linguistic structures[1,47], and the logical structures of semantically complex statements like concessives[37], biconditionals[37], conditionals[38], and disjunctions[17].

The visual world paradigm, compared with other psycholinguistic technique, also has some potential limitations. First, participants' interpretation of the spoken language is deduced from their eye movements on the visual world, not from the actual interpretation of the language stimuli per se. Henceforth, the language properties that can be studied with the visual world paradigm are limited to the ones that can be visualized, *i.e.*, they should be somehow related to the entities or events in the visual world. Second, the visual world paradigm used is normally more restricted than the actual visual world, with a limited set of pictured referents and a limited set of potential actions. This so-called closed-set problem[51] might create task-specific strategies that the observed language processing does not generalize beyond the specific situations created within the experiment. Henceforth, the paradigm might not be sensitive to characteristics of linguistic knowledge and experience lying outside of the closed-set that have been established on a given trial.

The visual world paradigm essentially explores the integration of the information from the visual domain and the information from the auditory domain. Theoretically, any information that can be processed by the two sensational domains can be potentially studied using this paradigm. For example, the visual world can be virtual reality or dynamic videos. The auditory input is not necessarily language, and can be format, such as music, sound from the natural world, *etc*. Furthermore, this paradigm can be extended further to explore the integration of the information from other domains, rather than the visual domain and the auditory domain. For example, researchers might use this technique to see how participants' fixations in the visual world are affected by different smells, different touches, *etc*.

ACKNOWLEDGMENTS:

This research was supported by Science Foundation of Beijing Language and Cultural University under the Fundamental Research Funds for the Central Universities (Approval number 15YJ050003).

DISCLOSURES:

The author declares that he has no competing financial interests.

REFERENCES:

1 Tanenhaus, M. K., Spivey-Knowlton, M. J., Eberhard, K. M. & Sedivy, J. C. Integration of visual and linguistic information in spoken language comprehension. *Science.* **268** (5217), 1632-1634, (1995).

2 Cooper, R. M. The control of eye fixation by the meaning of spoken language: A new methodology for the real-time investigation of speech perception, memory, and language processing. *Cognitive Psychology.* **6** (1), 84-107, (1974).

3 Salverda, A. P. & Tanenhaus, M. K. in *Research methods in psycholinguistics and the neurobiology of language: A practical guide* eds Annette M. B. de Groot & Peter Hagoort) (Wiley, 2017).

4 Huettig, F., Rommers, J. & Meyer, A. S. Using the visual world paradigm to study language processing: A review and critical evaluation. *Acta Psychologica.* **137** (2), 151-171, (2011).

5 Meyer, A. S., Sleiderink, A. M. & Levelt, W. J. M. Viewing and naming objects: Eye movements during noun phrase production. *Cognition.* **66** (2), B25-B33, (1998).

6 Griffin, Z. M. & Bock, K. What the eyes say about speaking. *Psychological Science.* **11** (4), 274-279, (2000).

7 Young, L. R. & Sheena, D. Survey of eye movement recording methods. *Behavior Research Methods & Instrumentation.* **7** (5), 397-429, (1975).

8 Conklin, K., Pellicer-Sánchez, A. & Carrol, G. *Eye-tracking: A guide for applied linguistics research*. (Cambridge University Press, 2018).

9 Duchowski, A. *Eye tracking methodology: Theory and practice*. 2 edn, (Springer, 2007).

10 Baayen, R. H., Davidson, D. J. & Bates, D. M. Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language.* **59** (4), 390-412, (2008).

11 Barr, D. J. Analyzing 'visual world' eyetracking data using multilevel logistic regression. *Journal of Memory and Language.* **59** (4), 457-474, (2008).

12 Nixon, J. S., van Rij, J., Mok, P., Baayen, R. H. & Chen, Y. The temporal dynamics of perceptual uncertainty: eye movement evidence from Cantonese segment and tone perception. *Journal of Memory and Language.* **90** 103-125, (2016).

13 Bolker, B. M. *et al.* Generalized linear mixed models: A practical guide for ecology and evolution. *Trends in Ecology and Evolution.* **24** (3), 127-135, (2009).

14 Mirman, D., Dixon, J. A. & Magnuson, J. S. Statistical and computational models of the visual world paradigm: Growth curves and individual differences. *Journal of Memory and Language.* **59** (4), 475-494, (2008).

15 Baayen, H., Vasishth, S., Kliegl, R. & Bates, D. The cave of shadows: Addressing the human factor with generalized additive mixed models. *Journal of Memory and Language.* **94** 206-234, (2017).

16 Baayen, R. H., van Rij, J., de Cat, C. & Wood, S. in *Mixed-Effects Regression Models in Linguistics* eds Dirk Speelman, Kris Heylen, & Dirk Geeraerts) Ch. 4, 49-69 (Springer, 2018).

17 Zhan, L. Scalar and ignorance inferences are both computed immediately upon encountering the sentential connective: The online processing of sentences with disjunction using the visual world paradigm. *Frontiers in Psychology.* **9**, (2018).

18 Maris, E. & Oostenveld, R. Nonparametric statistical testing of EEG- and MEG-data. *Journal of Neuroscience Methods.* **164** (1), 177-190, (2007).

19 Barr, D. J., Jackson, L. & Phillips, I. Using a voice to put a name to a face: The psycholinguistics of proper name comprehension. *Journal of Experimental Psychology-General.* **143** (1), 404-413, (2014).

20 Dahan, D., Tanenhaus, M. K. & Salverda, A. P. in *Eye movements: A window on mind and brain* eds Roger P. G. van Gompel, Martin H. Fischer, Wayne S. Murray, & And Robin L. Hill) Ch. 21, 471-486 (Elsevier, 2007).

21 Parkhurst, D., Law, K. & Niebur, E. Modeling the role of salience in the allocation of overt visual attention. *Vision Research.* **42** (1), 107-123, (2002).

22 Grice, H. P. in *Syntax and semantics* Vol. 3 *Speech Acts* eds Peter Cole & Jerry L. Morgan) 41-58 (Academic Press, 1975).

23 Sauerland, U. Scalar implicatures in complex sentences. *Linguistics and Philosophy.* **27** (3), 367-391, (2004).

24 Chierchia, G. Scalar implicatures and their interface with grammar. *Annual Review of Linguistics.* **3** (1), 245-264, (2017).

25 Fox, D. in *Presupposition and Implicature in Compositional Semantics* eds Uli Sauerland & Penka Stateva) Ch. 4, 71-120 (Palgrave Macmillan, 2007).

26 Meyer, M.-C. *Ignorance and grammar* Unpublished PhD Thesis thesis, Massachusetts Institute Of Technology, (2013).

27 SR Research Ltd. *EyeLink® 1000 Plus Technical Specifications* (Ottawa, Canada, 2017).

28 SR Research Ltd. *EyeLink-1000-Plus-Brochure* (2017).

29 McQueen, J. M. & Viebahn, M. C. Tracking recognition of spoken words by tracking looks to printed words. *The Quarterly Journal of Experimental Psychology.* **60** (5), 661-671, (2007).

30 Altmann, G. T. M. & Kamide, Y. Incremental interpretation at verbs: restricting the domain of subsequent reference. *Cognition.* **73** (3), 247-264, (1999).

31 Altmann, G. T. M. & Kamide, Y. The real-time mediation of visual attention by language and world knowledge: Linking anticipatory (and other) eye movements to linguistic processing. *Journal of Memory and Language.* **57** (4), 502-518, (2007).

32 Snedeker, J. & Trueswell, J. C. The developing constraints on parsing decisions: The role

of lexical-biases and referential scenes in child and adult sentence processing. *Cognitive Psychology.* **49** (3), 238-299, (2004).

33 Allopenna, P. D., Magnuson, J. S. & Tanenhaus, M. K. Tracking the time course of spoken word recognition using eye movements: Evidence for continuous mapping models. *Journal of Memory and Language.* **38** (4), 419-439, (1998).

34 Zhan, L., Crain, S. & Zhou, P. The online processing of only if and even if conditional statements: Implications for mental models. *Journal of Cognitive Psychology.* **27** (3), 367-379, (2015).

35 Zhan, L., Zhou, P. & Crain, S. Using the visual-world paradigm to explore the meaning of conditionals in natural language. *Language, Cognition and Neuroscience.* 10.1080/23273798.2018.1448935, (2018).

36 Brown-Schmidt, S. & Tanenhaus, M. K. Real-time investigation of referential domains in unscripted conversation: A targeted language game approach. *Cognitive Science.* **32** (4), 643-684, (2008).

37 Fernald, A., Pinto, J. P., Swingley, D., Weinberg, A. & McRoberts, G. W. Rapid gains in speed of verbal processing by infants in the 2nd year. *Psychological Science.* **9** (3), 228-231, (1998).

38 Trueswell, J. C., Sekerina, I., Hill, N. M. & Logrip, M. L. The kindergarten-path effect: studying on-line sentence processing in young children. *Cognition.* **73** (2), 89-134, (1999).

39 Zhou, P., Su, Y., Crain, S., Gao, L. Q. & Zhan, L. Children's use of phonological information in ambiguity resolution: a view from Mandarin Chinese. *Journal of Child Language.* **39** (4), 687-730, (2012).

40 Zhou, P., Crain, S. & Zhan, L. Grammatical aspect and event recognition in children's online sentence comprehension. *Cognition.* **133** (1), 262-276, (2014).

41 Zhou, P., Crain, S. & Zhan, L. Sometimes children are as good as adults: The pragmatic use of prosody in children's on-line sentence processing. *Journal of Memory and Language.* **67** (1), 149-164, (2012).

42 Moscati, V., Zhan, L. & Zhou, P. Children's on-line processing of epistemic modals. *Journal of Child Language.* **44** (5), 1025-1040, (2017).

43 Helfer, K. S. & Staub, A. Competing speech perception in older and younger adults: Behavioral and eye-movement evidence. *Ear and Hearing.* **35** (2), 161-170, (2014).

44 Dickey, M. W., Choy, J. W. J. & Thompson, C. K. Real-time comprehension of wh-movement in aphasia: Evidence from eyetracking while listening. *Brain and Language.* **100** (1), 1-22, (2007).

45 Magnuson, J. S. & Nusbaum, H. C. Acoustic differences, listener expectations, and the perceptual accommodation of talker variability. *Journal of Experimental Psychology-Human Perception and Performance.* **33** (2), 391-409, (2007).

46 Reinisch, E., Jesse, A. & McQueen, J. M. Early use of phonetic information in spoken word recognition: Lexical stress drives eye movements immediately. *Quarterly Journal of Experimental Psychology.* **63** (4), 772-783, (2010).

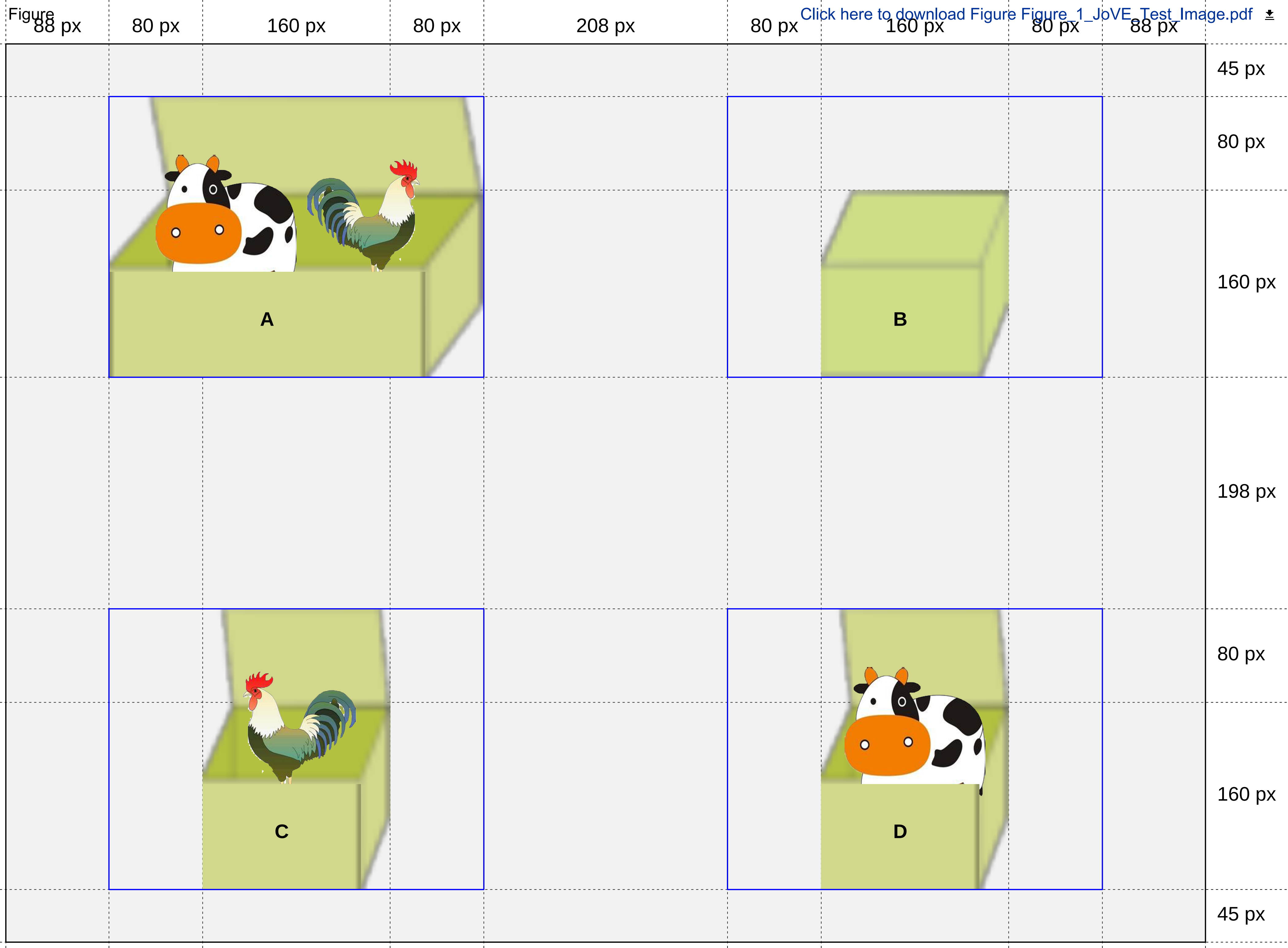
47 Chambers, C. G., Tanenhaus, M. K. & Magnuson, J. S. Actions and affordances in syntactic ambiguity resolution. *Journal of Experimental Psychology-Learning Memory and Cognition.* **30** (3), 687-696, (2004).

48 Tanenhaus, M. K. & Trueswell, J. C. in *Approaches to Studying World-Situated Language*

Use: Bridging the Language-as-Product and Language-as-Action Traditions eds John C. Trueswell & Michael K. Tanenhaus) (The MIT Press, 2005).

Figure

Click here to download Figure Figure_1_JoVE_Test_Image.pdf



Figure

a). **And**

小明的	箱子里	有	一只	奶牛	和	一只	公鸡
Xiaoming de	xiang zi li	you	yi zhi	nai niu	he	yi zhi	gong ji
Xiaoming's	box in	have	one-CL	cow	and	one-CL	rooster

Xiaoming's box contains a cow and a rooster.

b). **But**

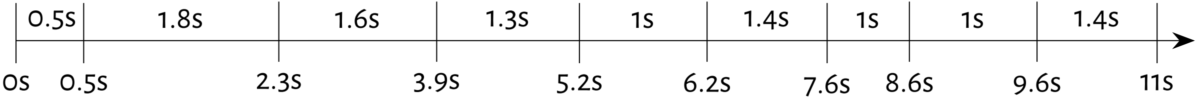
小明的	箱子里	有	一只	奶牛	但	没有	公鸡
Xiaoming de	xiangzi li	you	yi zhi	nai niu	dan	meiyou	gong ji
Xiaoming's	box in	have	one-CL	cow	but	not	rooster

Xiaoming's box contains a cow but not a rooster.

c). **Or**

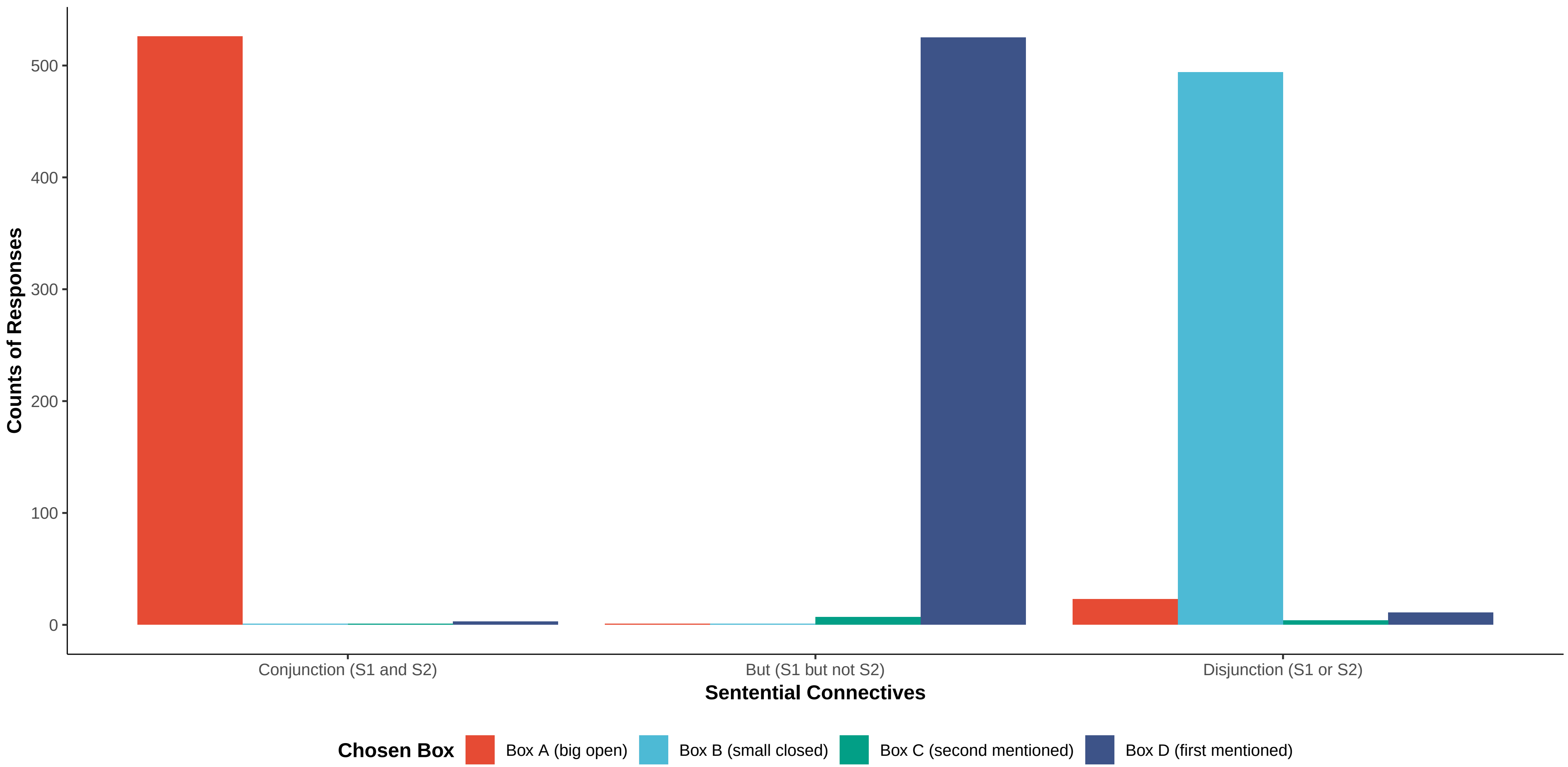
小明的	箱子里	有	一只	奶牛	或	一只	公鸡
Xiaoming de	xiang zi li	you	yi zhi	nainiu	huo	youzhi	gongji
Xiaoming's	box in	have	one-CL	cow	or	one-CL	rooster

Xiaoming's box contains a cow or a rooster.

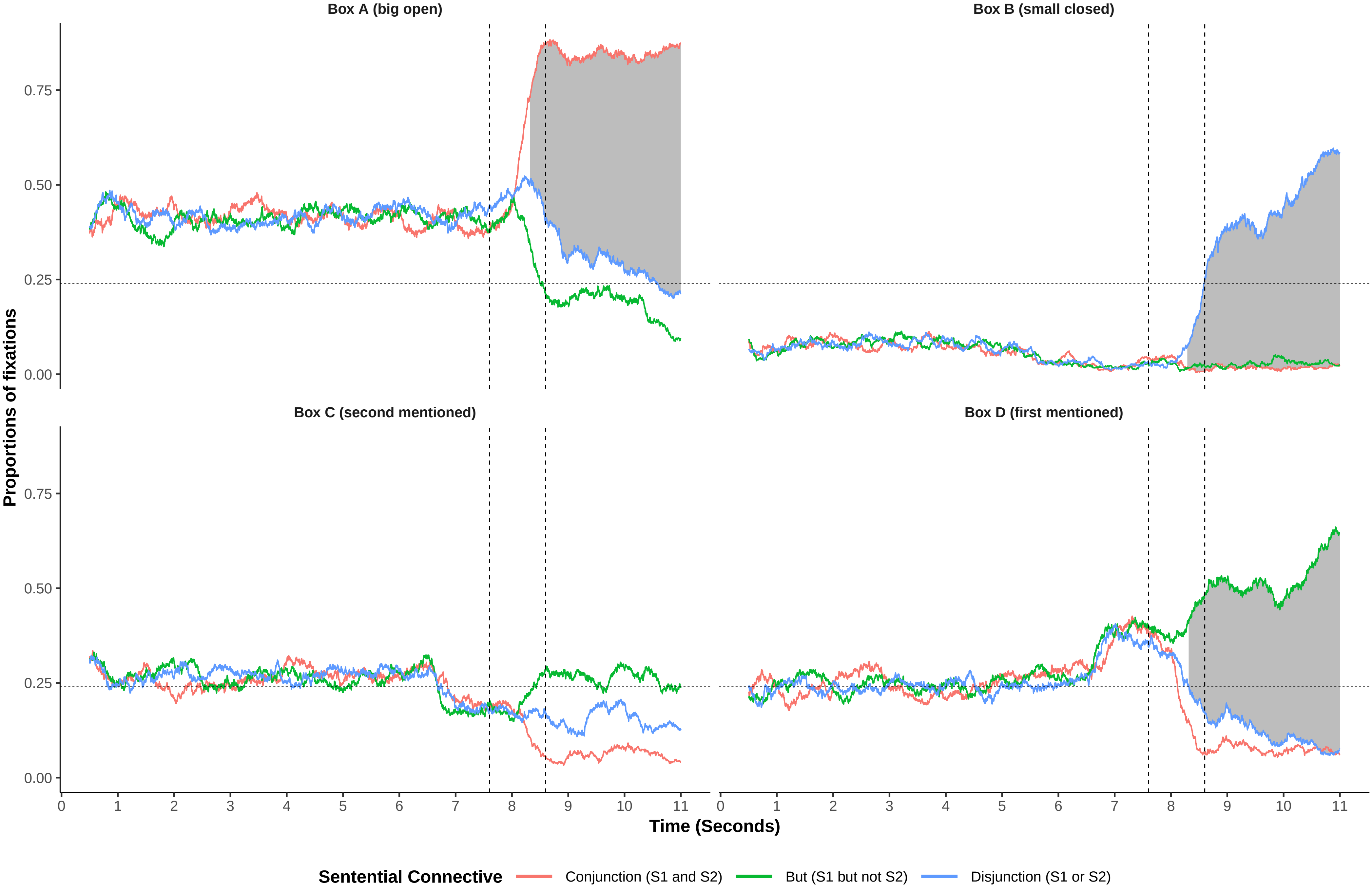


Figure

Click here to download Figure Figure_4_JoVE_Behavioral_responses.eps

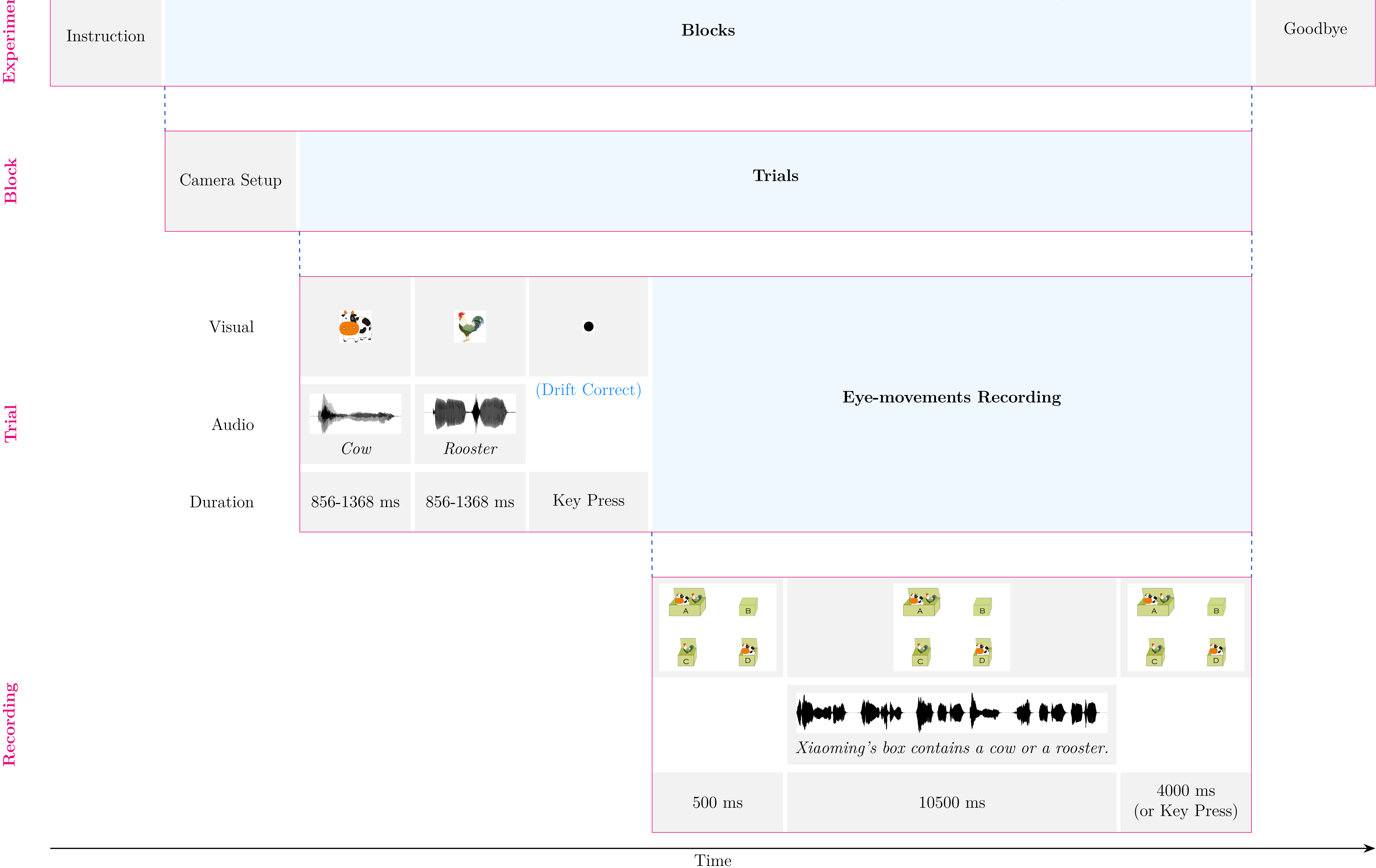


Click here to download Figure Figure_5_JoVE_Eye_Movements.eps



Figure

Click here to download Figure Figure_3_JoVE_ExperimentTimeLine.pdf



Name of Material/ Equipment	Company	Catalog Number
Pixelmator	Pixelmator Team	http://www.pixelmator.com/pro/
Praat	Open Sourse	http://www.fon.hum.uva.nl/praat/
Eyelink 1000plus	SR-Research, Inc	https://www.sr-research.com/products/eyelink-1000-plus/
Experimental Builder	SR-Research, Inc	https://www.sr-research.com/experiment-builder/
Data Viewer	SR-Research, Inc	https://www.sr-research.com/data-viewer/
R	Open Sourse	https://www.r-project.org

Comments/Description
image editing app
Sound analyses and editting software
remote infrared eye tracker
eye tracker software
eye tracker software
free software environment for statistical computing and graphics

JoVE JOURNAL OF VISUALIZED EXPERIMENTS
1 Alewife Center #200
Cambridge, MA 02140
tel. 617.945.9051
www.jove.com

ARTICLE AND VIDEO LICENSE AGREEMENT

Title of Article: Using eye movements in the visual world paradigm to explore the online processing of spoken language

Author(s): Likan Zhan

Item 1 (check one box): The Author elects to have the Materials be made available (as described at http://www.jove.com/author) via: ☑ Standard Access ☐ Open Access

Item 2 (check one box):

☑ The Author is NOT a United States government employee.

☐ The Author is a United States government employee and the Materials were prepared in the course of his or her duties as a United States government employee.

☐ The Author is a United States government employee but the Materials were NOT prepared in the course of his or her duties as a United States government employee.

ARTICLE AND VIDEO LICENSE AGREEMENT

1. Defined Terms. As used in this Article and Video License Agreement, the following terms shall have the following meanings: "**Agreement**" means this Article and Video License Agreement; "**Article**" means the article specified on the last page of this Agreement, including any associated materials such as texts, figures, tables, artwork, abstracts, or summaries contained therein; "**Author**" means the author who is a signatory to this Agreement; "**Collective Work**" means a work, such as a periodical issue, anthology or encyclopedia, in which the Materials in their entirety in unmodified form, along with a number of other contributions, constituting separate and independent works in themselves, are assembled into a collective whole; "**CRC License**" means the Creative Commons Attribution-Non Commercial-No Derivs 3.0 Unported Agreement, the terms and conditions of which can be found at: http://creativecommons.org/licenses/by-nc-nd/3.0/legalcode; "**Derivative Work**" means a work based upon the Materials or upon the Materials and other pre-existing works, such as a translation, musical arrangement, dramatization, fictionalization, motion picture version, sound recording, art reproduction, abridgment, condensation, or any other form in which the Materials may be recast, transformed, or adapted; "**Institution**" means the institution, listed on the last page of this Agreement, by which the Author was employed at the time of the creation of the Materials; "**JoVE**" means MyJove Corporation, a Massachusetts corporation and the publisher of *The Journal of Visualized Experiments;* "**Materials**" means the Article and / or the Video; "**Parties**" means the Author and JoVE; "**Video**" means any video(s) made by the Author, alone or in conjunction with any other parties, or by JoVE or its affiliates or agents, individually or in collaboration with the Author or any other parties, incorporating all or any portion of the Article, and in which the Author may or may not appear.

2. Background. The Author, who is the author of the Article, in order to ensure the dissemination and protection of the Article, desires to have the JoVE publish the Article and create and transmit videos based on the Article. In furtherance of such goals, the Parties desire to memorialize in this Agreement the respective rights of each Party in and to the Article and the Video.

3. Grant of Rights in Article. In consideration of JoVE agreeing to publish the Article, the Author hereby grants to JoVE, subject to **Sections 4** and **7** below, the exclusive, royalty-free, perpetual (for the full term of copyright in the Article, including any extensions thereto) license (a) to publish, reproduce, distribute, display and store the Article in all forms, formats and media whether now known or hereafter developed (including without limitation in print, digital and electronic form) throughout the world, (b) to translate the Article into other languages, create adaptations, summaries or extracts of the Article or other Derivative Works (including, without limitation, the Video) or Collective Works based on all or any portion of the Article and exercise all of the rights set forth in (a) above in such translations, adaptations, summaries, extracts, Derivative Works or Collective Works and (c) to license others to do any or all of the above. The foregoing rights may be exercised in all media and formats, whether now known or hereafter devised, and include the right to make such modifications as are technically necessary to exercise the rights in other media and formats. If the "Open Access" box has been checked in **Item 1** above, JoVE and the Author hereby grant to the public all such rights in the Article as provided in, but subject to all limitations and requirements set forth in, the CRC License.

4. Retention of Rights in Article. Notwithstanding the exclusive license granted to JoVE in **Section 3** above, the Author shall, with respect to the Article, retain the non-exclusive right to use all or part of the Article for the non-commercial purpose of giving lectures, presentations or teaching classes, and to post a copy of the Article on the Institution's website or the Author's personal website, in each case provided that a link to the Article on the JoVE website is provided and notice of JoVE's copyright in the Article is included. All non-copyright intellectual property rights in and to the Article, such as patent rights, shall remain with the Author.

5. Grant of Rights in Video – Standard Access. This **Section 5** applies if the "Standard Access" box has been checked in **Item 1** above or if no box has been checked in **Item 1** above. In consideration of JoVE agreeing to produce, display or otherwise assist with the Video, the Author hereby acknowledges and agrees that, Subject to **Section 7** below, JoVE is and shall be the sole and exclusive owner of all rights of any nature, including, without limitation, all copyrights, in and to the Video. To the extent that, by law, the Author is deemed, now or at any time in the future, to have any rights of any nature in or to the Video, the Author hereby disclaims all such rights and transfers all such rights to JoVE.

6. Grant of Rights in Video – Open Access. This **Section 6** applies only if the "Open Access" box has been checked in **Item 1** above. In consideration of JoVE agreeing to produce, display or otherwise assist with the Video, the Author hereby grants to JoVE, subject to **Section 7** below, the exclusive, royalty-free, perpetual (for the full term of copyright in the Article, including any extensions thereto) license (a) to publish, reproduce, distribute, display and store the Video in all forms, formats and media whether now known or hereafter developed (including without limitation in print, digital and electronic form) throughout the world, (b) to translate the Video into other languages, create adaptations, summaries or extracts of the Video or other Derivative Works or Collective Works based on all or any portion of the Video and exercise all of the rights set forth in (a) above in such translations, adaptations, summaries, extracts, Derivative Works or Collective Works and (c) to license others to do any or all of the above. The foregoing rights may be exercised in all media and formats, whether now known or hereafter devised, and include the right to make such modifications as are technically necessary to exercise the rights in other media and formats. For any Video to which this Section 6 is applicable, JoVE and the Author hereby grant to the public all such rights in the Video as provided in, but subject to all limitations and requirements set forth in, the CRC License.

7. Government Employees. If the Author is a United States government employee and the Article was prepared in the course of his or her duties as a United States government employee, as indicated in **Item 2** above, and any of the licenses or grants granted by the Author hereunder exceed the scope of the 17 U.S.C. 403, then the rights granted hereunder shall be limited to the maximum rights permitted under such statute. In such case, all provisions contained herein that are not in conflict with such statute shall remain in full force and effect, and all provisions contained herein that do so conflict shall be deemed to be amended so as to provide to JoVE the maximum rights permissible within such statute.

8. Likeness, Privacy, Personality. The Author hereby grants JoVE the right to use the Author's name, voice, likeness, picture, photograph, image, biography and performance in any way, commercial or otherwise, in connection with the Materials and the sale, promotion and distribution thereof. The Author hereby waives any and all rights he or she may have, relating to his or her appearance in the Video or otherwise relating to the Materials, under all applicable privacy, likeness, personality or similar laws.

9. Author Warranties. The Author represents and warrants that the Article is original, that it has not been published, that the copyright interest is owned by the Author (or, if more than one author is listed at the beginning of this Agreement, by such authors collectively) and has not been assigned, licensed, or otherwise transferred to any other party. The Author represents and warrants that the author(s) listed at the top of this Agreement are the only authors of the Materials. If more than one author is listed at the top of this Agreement and if any such author has not entered into a separate Article and Video License Agreement with JoVE relating to the Materials, the Author represents and warrants that the Author has been authorized by each of the other such authors to execute this Agreement on his or her behalf and to bind him or her with respect to the terms of this Agreement as if each of them had been a party hereto as an Author. The Author warrants that the use, reproduction, distribution, public or private performance or display, and/or modification of all or any portion of the Materials does not and will not violate, infringe and/or misappropriate the patent, trademark, intellectual property or other rights of any third party. The Author represents and warrants that it has and will continue to comply with all government, institutional and other regulations, including, without limitation all institutional, laboratory, hospital, ethical, human and animal treatment, privacy, and all other rules, regulations, laws, procedures or guidelines, applicable to the Materials, and that all research involving human and animal subjects has been approved by the Author's relevant institutional review board.

10. JoVE Discretion. If the Author requests the assistance of JoVE in producing the Video in the Author's facility, the Author shall ensure that the presence of JoVE employees, agents or independent contractors is in accordance with the relevant regulations of the Author's institution. If more than one author is listed at the beginning of this Agreement, JoVE may, in its sole discretion, elect not take any action with respect to the Article until such time as it has received complete, executed Article and Video License Agreements from each such author. JoVE reserves the right, in its absolute and sole discretion and without giving any reason therefore, to accept or decline any work submitted to JoVE. JoVE and its employees, agents and independent contractors shall have

1 Alewife Center #200
Cambridge, MA 02140
tel. 617.945.9051
www.jove.com

ARTICLE AND VIDEO LICENSE AGREEMENT

full, unfettered access to the facilities of the Author or of the Author's institution as necessary to make the Video, whether actually published or not. JoVE has sole discretion as to the method of making and publishing the Materials, including, without limitation, to all decisions regarding editing, lighting, filming, timing of publication, if any, length, quality, content and the like.

11. Indemnification. The Author agrees to indemnify JoVE and/or its successors and assigns from and against any and all claims, costs, and expenses, including attorney's fees, arising out of any breach of any warranty or other representations contained herein. The Author further agrees to indemnify and hold harmless JoVE from and against any and all claims, costs, and expenses, including attorney's fees, resulting from the breach by the Author of any representation or warranty contained herein or from allegations or instances of violation of intellectual property rights, damage to the Author's or the Author's institution's facilities, fraud, libel, defamation, research, equipment, experiments, property damage, personal injury, violations of institutional, laboratory, hospital, ethical, human and animal treatment, privacy or other rules, regulations, laws, procedures or guidelines, liabilities and other losses or damages related in any way to the submission of work to JoVE, making of videos by JoVE, or publication in *JoVE* or elsewhere by JoVE. The Author shall be responsible for, and shall hold JoVE harmless from, damages caused by lack of sterilization, lack of cleanliness or by contamination due to the making of a video by JoVE its employees, agents or independent contractors. All sterilization, cleanliness or decontamination procedures shall be solely the responsibility of the Author and shall be undertaken at the Author's expense. All indemnifications provided herein shall include JoVE's attorney's fees and costs related to said losses or damages. Such indemnification and holding harmless shall include such losses or damages incurred by, or in connection with, acts or omissions of JoVE, its employees, agents or independent contractors.

12. Fees. To cover the cost incurred for publication, JoVE must receive payment before production and publication the Materials. Payment is due in 21 days of invoice. Should the Materials not be published due to an editorial or production decision, these funds will be returned to the Author. Withdrawal by the Author of any submitted Materials after final peer review approval will result in a US$1,200 fee to cover pre-production expenses incurred by JoVE. If payment is not received by the completion of filming, production and publication of the Materials will be suspended until payment is received.

13. Transfer, Governing Law. This Agreement may be assigned by JoVE and shall inure to the benefits of any of JoVE's successors and assignees. This Agreement shall be governed and construed by the internal laws of the Commonwealth of Massachusetts without giving effect to any conflict of law provision thereunder. This Agreement may be executed in counterparts, each of which shall be deemed an original, but all of which together shall be deemed to me one and the same agreement. A signed copy of this Agreement delivered by facsimile, e-mail or other means of electronic transmission shall be deemed to have the same legal effect as delivery of an original signed copy of this Agreement.

A signed copy of this document must be sent with all new submissions. Only one Agreement required per submission.

CORRESPONDING AUTHOR:

Name:	Likan Zhan
Department:	Institute for Speech Pathology and the Brain Science
Institution:	Beijing Language and Culture University
Article Title:	Using eye movements in the visual world paradigm to explore the online processing of spoken language
Signature:	战立侃
Date:	2018.03.01

Please submit a signed and dated copy of this license by one of the following three methods:

1) Upload a scanned copy of the document as a pfd on the JoVE submission site;
2) Fax the document to +1.866.381.2236;
3) Mail the document to JoVE / Attn: JoVE Editorial / 1 Alewife Center #200 / Cambridge, MA 02139

For questions, please email submissions@jove.com or call +1.617.945.9051

Dear Professor Bing Wu,

I really appreciate the detailed comments and the valuable suggestions from the editorial committee and the four reviewers, and thank you for giving us the opportunity to revise my manuscript: "*Using eye movements recorded in the visual world paradigm to explore the online processing of spoken language*" (Manuscript number: 58086). I have thoroughly revised the manuscript to address all the concerns put forward by editorial board. The following paragraphs summarized our responses to the editorial committee and each of the reviewers' comments.

1. For steps that are done using software (comments 3, 4, and 6), I added a step-wise description of every software usage. To make the description clearer, I split the experimental procedure section in the previous version into two different sections in the current version: Build the experimental script (section 3) and conduct the experiment (section 5). I also redrew Figure 3 to reflect these changes. The protocol now is too long, so I highlighted some steps in yellow to indicate that these steps should be featured in the video.
2. As the editorial board suggested (comments 1, 2), I have removed the personal pronouns and commercial language.
3. I have also explained more explicitly our standard to define "normal or correct normal vision" (comment 5): *As a general guideline, a participant is regarded as eligible as long as the participant can see the test images at a distance of about 60 centimeters* (lines 543-544).

This summarizes how we have responded to the thoughtful comments from the three reviewers and the editorial comments. All the changes, including the ones that are not mentioned in this rebuttal letter, are tracked and highlighted in the revised manuscript. I appreciate the careful readings the editorial board gave to the paper. I have tried to address the comments and criticisms in detail, and I think that this has resulted in a stronger paper.

Kind regards,

Likan Zhan

From: **Frontiers in Psychology** psychology@frontiersin.org
Subject: Re: Reprint of the images used in an published article
Date: April 24, 2018 at 7:59 PM
To: likanzhan@icloud.com

FP

Dear Dr. Zhan,

Thank you for your message. Please note that Frontiers does not own the copyright to the article you are referring to. All articles published at Frontiers are under a CC-BY license, meaning that authors retain the full copyright to their work.
Therefore, you have complete liberty to use the images and/or tables in your article.

I remain available should you have any further question.

Warm Regards,

Chloe Blundell
Journal Development Specialist

Frontiers | Editorial Office - Journal Development Team
Journal Development Manager: Adriana Timperi

Frontiers
www.frontiersin.org
Avenue du Tribunal Fédéral 34
Lausanne, Switzerland
Office T +41(0)21 510 17 11

On 24 April 2018 at 13:51, Frontiers in Psychology <psychology@frontiersin.org> wrote:

Dear Dr. Zhan,

Thank you for your message. Please note that Frontiers does not own the copyright to the article you are referring to. All articles published at Frontiers are under a CC-BY license, meaning that authors retain the full copyright to their work.
Therefore, you have complete liberty to use the images and/or tables in your article.

I remain available should you have any further question.

Warm Regards,

Chloe Blundell
Journal Development Specialist

Frontiers | Editorial Office - Journal Development Team
Journal Development Manager: Adriana Timperi

Frontiers
www.frontiersin.org
Avenue du Tribunal Fédéral 34
Lausanne, Switzerland
Office T +41(0)21 510 17 11

Frontiers journals lead in citations in their fields and rank in the top Impact Factor percentiles. Read the full analysis here.

For technical issues, please visit our Frontiers Help Center https://zendesk.frontiersin.org/hc/en-us

On 24 April 2018 at 13:08, Frontiers Editorial Office <editorial.office@frontiersin.org> wrote:

---------- Forwarded message ----------
From: 战立侃 <likanzhan@icloud.com>
Date: Tue, Apr 24, 2018 at 10:30 AM
Subject: Reprint of the images used in an published article
To: editorial.office@frontiersin.org

Dear editorial board of Frontiers,

I recently published a paper on Frontiers in Psychology:

https://www.frontiersin.org/articles/10.3389/fpsyg.2018.00061/full

And now I'm intending to reprint some of the images and tables used in that article (for none commercial purposes).

Is there anything I need to do obtain the copyright or something?

Thanks

Likan